Dean&Francis

ISSN 2959-6157

Research and Analysis of Railway Obstacle Detection Technology Based on Deep Learning

Zhaoyun Chen

College of Computer Engineering, Jiangsu Second Normal University, Nanjing, China
zhaoyunchen33@gmail.com

Abstract:

The advancement of intelligent detection technologies has made highly reliable obstacle identification a critical factor in enhancing the operational safety of railways. Nevertheless, traditional detection methods (such as rail-side infrared counters and vehicle-mounted millimeter-wave radars) are limited by inherent defects such as strong environmental sensitivity, high detection rate of small targets and no classification ability, making it difficult to meet the safety needs in complex scenarios. In recent years, deep learning has promoted the transition of railway obstacle detection technology to intelligence with its multimodal feature autonomous extraction and dynamic scene modeling capabilities. This review offers a thorough analysis of the historical development and advancements in railway obstacle detection methods, highlighting the transformative advances brought about by deep learning-based models, comprehensively compares the application characteristics and scenarios of various technologies, and discusses the application and evolution of deep learning in the two core tasks of target recognition and semantic segmentation, covering representative models such as convolutional neural network and Transformer. This article also uses the analysis of public data sets to help the new generation of railway obstacle detection technology to provide a theoretical basis and also facilitate researchers to quickly carry out experimental verification.

Keywords: Deep Learning; Obstacle Detection; Convolutional Neural Networks; Rail Transit.

1. Introduction

Since 2012, In the field of railway safety monitoring, deep learning and deep empowerment of computer vision are triggering a series of profound technological innovations. The intelligent detection system of rail obstacles has become a hot topic in research due to its key role in ensuring the safety of train operations. At present, in the field of transportation, a variety of technical paths have been developed for railway obstacle detection, mainly covering computer vision-based, radar (millimeter wave radar) as the core, and vision and lidar (light detection and ranging) fusion detection system design [1]. Compared with traditional rules-basedor manual features, deep learning, as a machine learning technology, has powerful feature extraction and classification capabilities, providing new ideas for obstacle detection. Convolutional neural network, recurrent neural network and generative adversarial network, as typical deep learning architectures, each have unique advantages and are suitable for specific scenarios. With the optimization of model structure in recent years and the improvement of training algorithms, the accuracy and robustness of detection have been significantly improved. However, due to the particularity of railway scenarios, railway obstacle detection technology still faces severe challenges in practical applications. The main contributions of this paper are: (1) A systematic review of the development of railway obstacle detection technology based on deep learning, and an in-depth discussion of the application and evolution of deep learning in the two core tasks of target recognition and semantic segmentation. (2) By introducing and evaluating multiple public data sets suitable for this field, it provides convenience for model training (3) Summarizing the current shortcomings of railway obstacle detection technology and looking forward to future improvement directions.

2. Technology Evolution of Railway Obstacle Detection

Key technologies based on deep learning obstacle detection include three categories: data enhancement technology, multi-scale feature fusion and object detection algorithm. The following article will review the technological evolution of railway obstacle detection since 2010. Figure 1 is the timeline of the development of obstacle detection technology, marking a milestone breakthrough at a critical time node.

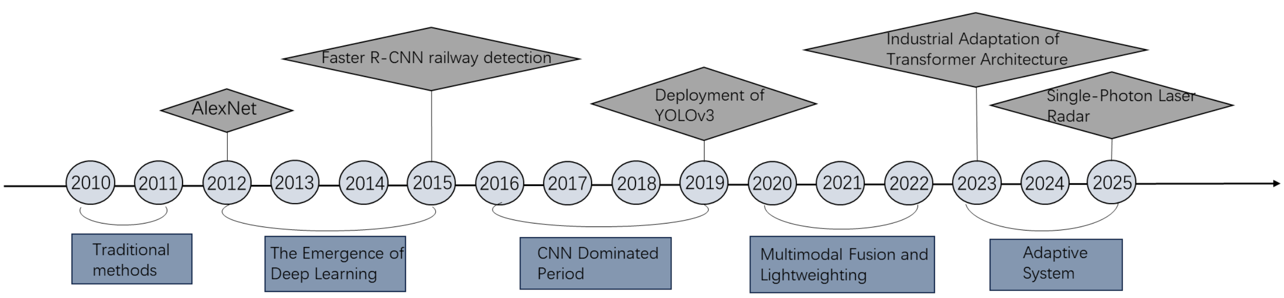

Fig. 1 Timeline diagram of the development of obstacle detection technology

2.1 Traditional Computer Vision (2010-2011)

Traditional railway obstacle detection methods are mainly based on image processing and computer vision technologies, such as edge detection, region growth, feature matching, etc. However, the detection effect in complex environments is poor and it is difficult to meet actual needs. For example, the OpenCV algorithm has a significant performance gap with the deep learning method [2]. On the same test set (RailSem19), the mAP of the traditional method is 58.2% (OpenCV), which is far lower than the 92.7% (YOLOv5+Transformer) of the deep learning method.

2.2 The introduction and development stage of deep learning (2012-2019)

In 2012, AlexNet was the champion model of the ImageNet competition, and this breakthrough marked a turning point, establishing deep convolutional neural networks as the leading paradigm for computer vision tasks. It was not until 2013 that deep learning began to show great potential in image classification tasks. With the improvement and development of convolutional neural network models, it gradually replaced traditional manual feature extraction. In 2016, FasterR-CNN (RegionProposal Network, RPN) proposed [3], which has end-to-end training capabilities, greatly improving detection accuracy and speed. Since its birth in 2015 [4], the YOLO series has completely changed the field of target detection with its efficient characteristics of „single reasoning“. YOLO 3 was proposed in 2018 [5] and was officially deployed in railway obstacle detection in 2019.

2.3 Fusion of deep learning and multimodality (2020-2022)

2020 to 2022 is a period of rapid development of deep learning technology, at this stage, multimodal fusion approaches are increasingly being implemented in the context of railway obstacle identification. In 2021, LiDAR

ISSN 2959-6157

point cloud fusion, the Cui team published an open source project fusion_pointclouds. After calibration of unmanned vehicle multi-lidar in the Ubuntu environment, multiple LiDAR point cloud topics/coordinate systems are fused into a ros point cloud topic through PCL (PointCloudLibrary) to facilitate later point cloud ground segmentation and ground processing [6].

In 2022, Transformer architecture applied to railway obstacle detection, combining lidar and camera image data for fusion, significantly improving the accuracy and robustness of detection under poor image conditions such as poor illumination and sensor misalignment [7].

2.4 Deep reinforcement learning and adaptive model (2023-to-present)

The practical turning point in 2023, the dual-mode (visible light + infrared) system is piloted on high-speed rail. In 2025, the Chinese Academy of Aerospace and Space Institute completed a single-photon lidar system test, using SPAD array to achieve 2km ultra-long-distance imaging, with a resolution of up to centimeters, breaking through the bottleneck of light-free environment detection in the tunnel. At present, deep reinforcement learning (DRL) has begun to be applied to obstacle detection and decision-making in dynamic environments to help the system continuously optimize detection accuracy. In the railway system, self-supervised learning is applied to unsupervised data learning in the event of scarcity of labeled data to further improve the adaptability of the model in unknown environments.

3. Core Technology Methods

Deep learning-based obstacle detection has emerged as a pivotal research area within environmental perception. The principal architectures employed in this domain primarily include convolutional neural networks, recurrent neural networks, and more recently, attention-based models such as the Transformer. CNNs demonstrate considerable strengths in image feature extraction and object localization, owing to their properties of local receptive fields and parameter sharing. RNNs, with their capacity for temporal modeling, are frequently applied to analyze dynamic obstacles across sequential frames or multimodal sensor data. Meanwhile, attention-driven frameworks enhance detection precision and robustness for small or partially occluded objects in complex environments by emphasizing relevant regions. Today's railway obstacle detection technology is undergoing the evolution of lightweight → multimodal → large model. Target detection approaches industrial threshold in real time, and semantic segmentation has made breakthroughs in specialization accuracy and cross-scene generalization. Among them, YOLOv9-CMDA technology is used in vehicle-mounted real-time braking [8], which greatly reduces accident losses; MaskTransfiner's accuracy in track micro-defect detection is as high as 96.8% [9], effectively reducing manual inspection costs.

3.1 The leading method of object detection

The target detection task focuses on real-time positioning of track invasions (such as pedestrians, rockfalls, vehicles), and requires balance accuracy, speed and small target adaptability. The advancement of convolutional neural networks has led to the widespread adoption of both single-stage detectors and two-stage detectors in the domain of object recognition[10]. At the same time, some improved algorithms based on YOLO are emerging. These algorithms have conducted in-depth research and optimization in terms of model lightweighting, multi-scale detection, data enhancement strategies, training strategy optimization, etc., and have achieved significant performance improvements.

3.1.1 Lightweight YOLO series improvements

To address issues such as excessive parameters, low detection accuracy, slow inference speed, and limited generalization capability, HuangfuJY et al. developed a real-time traffic detection model based on YOLOv5, incorporating GhostNet and an attention mechanism. A genetic algorithm-enhanced K-means clustering approach was employed to derive optimized anchor boxes tailored for vehicle detection. Lightweight Ghost convolutions were adopted for feature extraction, and a C3Ghost module structured around CSP was designed to significantly reduce parameter count and computational overhead while accelerating processing. Furthermore, Transformer blocks and the CBAM attention mechanism were incorporated into the feature fusion layer to augments the model's feature representation capability and improve focus on regions of interest in dense traffic scenarios [11].

3.1.2 Transformer-Convolution Hybrid Architecture

The AE-Net model was proposed by ZhaoZ team and an enhanced and lightweight Transformer module (ETM) was built to enhance the global modeling capabilities of the model. A lightweight feature integration module (LIM) is introduced to consolidate multi-branch feature representations while minimizing model complexity. Additionally, a refined multi-scale feature fusion module (RFM) is employed to improve detection performance for objects at various scales. This algorithm achieves 95.29%mAP and 145 frames per second on the railway dataset, which is better than YOLOv5 [12]. Especially suitable for small object detection.

3.2 The dominant method of image semantic segmentation

Image semantic segmentation refers to assigning each pixel in an image to its corresponding semantic category, thereby achieving fine-grained understanding and analysis of the image. The image semantic segmentation algorithm of deep learning is mainly designed based on the architecture of Convolutional Neural Networks (CNN). These algorithms enable the network to learn the feature representation of different objects in the image by using a large amount of annotated image data during the training phase.

3.2.1 Fine segmentation of orbital structures

The approach introduced by the LiR team, termed DynaMask, employs a dual-stage feature pyramid architecture equipped with adaptive feature integration mechanisms. This design progressively refines the resolution of mask predictions, thereby enhancing the quality of instance segmentation. Furthermore, to mitigate the growing computational and memory demands caused by processing high-resolution masks, the method incorporates a lightweight Mask Switch Module (MSM). This component dynamically determines the most suitable mask resolution for each instance individually, maintaining high segmentation precision with only a marginal increase in computational cost [13].

3.2.2 Generalization segmentation of visual big model

Shanghai Artificial Intelligence Laboratory, Tsinghua University and other institutions jointly developed InternImage model, which uses a new deformable convolution (DCNv3) as the core operator, and correspondingly combines the advantages of visual transformer (ViT) and convolutional neural network (CNN), showing strong performance and flexibility [14]. And in 2025, the latest invention patent "Railway Unmanned Driving Obstacle Detection and Positioning Method, Device, Equipment and Media" obtained by the Hi-Tech University of Technology demonstrates a technical path combining point cloud data processing and deep learning, which can detect and locate known and unknown obstacles at the same time [15].

4. Introduction to Railway Obstacle Detection Data

This section summarizes multiple open source data sets commonly used in the field of obstacle detection in railway scenarios. Table 1 summarizes the modalities and scenarios of each data set, covering three sources: real operation lines, test fields and synthetic data, providing researchers with the convenience of algorithm training and performance verification. The data set is designed for different detection tasks (static target recognition, dynamic intrusion warning, multimodal fusion), with both scene diversity and physical authenticity.

Table 1. Railway obstacle detection data set

Dataset	Modal	Scene
OSDaR23	RGB+IR+LiDAR	Real operating environment and multiple climatic conditions
SynDRA-BBox	RGB+Depth	Physical real extreme scenes
Rail-DB	RGB	Rail reflection and complex background

The OSDaR23 multimodal dataset is a 21-sequence multi-sensor dataset released by the Technology University of Hamburg, Germany, and captured in September 2021. It is the first railway obstacle detection dataset that synchronizes visible light, infrared and lidar. In addition to the original data, this dataset also contains 204091 polylines, polygons, rectangles and cuboid notes for 20 different object classes [16]. OSDaR23 has developed a CV-based collision prediction system for automatic train operation (ATO), and also provides special cases such as thermal diffusion effect in the tunnel, rain fog point cloud decay, etc. Taking this dataset as an example, some dataset samples are given, as shown in Figure 2.

SynDRA-BBox synthetic dataset was released by the University of Santa Ana, Italy, and the first physical-level simulation dataset for railway 3D detection. The public dataset includes camera, depth and LiDAR data, coupled with a variety of annotations to support the evaluation of visual algorithms in a railway environment. SynDRA-BBox is the first publicly available synthetic dataset that facilitates both 2D and 3D object detection in addition to semantic segmentation in the transportation field. The dataset includes seven different railway intersection scenarios, as well as additional scenarios in a large railway station environment. The dataset also provides high-resolution RGB images and synthetic LiDAR point clouds, as well as detailed annotations in the image and point cloud domains [17]. SynDRA-BBox aims to solve the academic

ISSN 2959-6157

challenge of scarce real labeling data in railway automation research.

Rail-DB track structure data set The railway detection data set created by Shenzhen University of Technology, contains 7432 pairs of images and their annotations, covering a variety of lighting, road structures and viewing angle conditions, and pixel-level analysis of track areas and safety boundaries. Rail-DB is the first public railway detection data set, filling the gap in the lack of public benchmarks in this field, and proposes an efficient row-based track detection method, Rail-Net, which includes a lightweight convolutional backbone network and anchor classifier [18], which promotes the diversification and efficient development of railway detection technology.

Fig. 2 Sample OSDaR23 dataset

5. Summary and Prospect of Existing Railway Intelligent Obstacle Detection Technology

In recent years, the domain of railway obstacle detection has witnessed substantial advances, with deep learning-based methods achieving considerable improvements in critical areas including object detection and semantic segmentation. Through systematic research, it can be found that existing research mainly focuses on the direction of multimodal robustness, synthetic data-driven, and structural specialization. However, due to the complexity of computing and environmental dynamics, the current railway obstacle detection technology still faces many challenges in practical applications. Based on existing research results, this paper systematically summarizes and prospects the existing problems and future development directions of railway obstacle detection.

5.1 Key issues currently available

The algorithm faces a trade-off between real-time performance and accuracy, with inference latency on edge devices exceeding permissible thresholds. Multimodal inputs (including visible light, infrared, and semantic maps) lead to a substantial increase in feature channels, resulting in high computational redundancy in conventional convolution operations. Furthermore, lightweight models exhibit limited feature representation capacity. The Rail-DB system requires millimeter-level detection of track components and high-resolution feature extraction, which doubles computational demands. Consequently, memory bandwidth on edge devices becomes a critical bottleneck.

Small sample learning limitations. Due to the limited sample size and uneven distribution among classes, some key obstacles appear very low frequency. Although existing methods alleviate this problem through data augmentation and transfer learning, there are still obvious limitations in zero-sample obstacle recognition and open set detection.

5.2 Future development direction

Under the constraints of the above challenges, there are still significant shortcomings in the application effect, real-time reasoning capabilities and data utilization efficiency of existing railway obstacle detection technology in complex environments, and further optimization and breakthroughs are needed. Looking to the future, in response to the current challenges, subsequent research can be carried out from the following three directions:

Coordinated optimization of algorithm efficiency and accuracy. Develop a new lightweight architecture design, develop a dedicated neural architecture search (NAS) framework for railway scenarios, and automatically generate an optimal model that meets delay constraints; explore dynamic sparse convolution technology, adaptively adjust the calculation path according to the complexity of the input scenario; build a multi-scale feature distillation mechanism to retain high-resolution detailed features in the lightweight model.

Efficient fusion of multimodal data. Perform cross-modal feature compression, design channel-space dual-dimensional attention mechanism, dynamically suppress redundant modality; develop multimodal knowledge distillation technology to compress multi-sensor information to a single-modal network. Establish a heterogeneous graph neural network, build a 3D point cloud-image cross-modal graph, and achieve feature alignment through graph attention (GAT); propose a time-sequence multimodal Transformer, and use the self-attention mechanism to model the sensor space-time correlation.

Breakthrough in small sample and open set detection. Through meta-learning and incremental learning, a meta-learning benchmark for railway scenarios is constructed to support 5-shot obstacle recognition. The PROOF jointly proposed by Nanjing University and Nanyang Technological Technology can enable pre-trained vision-language models to continuously learn new categories and overcome catastrophic forgetting [19], but whether they are applied to the actual railway environment remains to be studied.

6. Conclusion

This paper systematically sorts out the development history of railway obstacle detection technology, summarizes and analyzes the latest achievements, and raises future challenges. Although the utilization of deep learning frameworks has markedly improved the performance of object detection and semantic segmentation methods in the context of railway obstacle identification, it still faces challenges in complex environment robustness, edge real-timeness and data efficiency. Future research needs to be carried out around the directions of multi-task collaborative optimization, cross-domain adaptive learning, hardware-algorithm collaborative optimization. Only by overcoming the three major challenges of environmental dynamics, tight computing constraints, and data sparseness can we promote the evolution of railway obstacle detection technology toward high accuracy, low latency and strong generalization, and provide core technical support for the next generation of intelligent railway active safety protection systems.

References

[1] Zhang Q, Yan F, Song W, Wang R, Li G. Automatic Obstacle Detection Method for the Train Based on Deep Learning. Sustainability. 2023, 15, 1184.

[2] Bradski G. The opencv library. Dr. Dobb's Journal: Software Tools for the Professional Programmer. 2000, 25(11): 120-123.

[3] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence. 2016, 39(6): 1137-1149.

[4] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, 779-788.

[5] Redmon J, Farhadi A. Yolov3: An incremental improvement. arxiv preprint arxiv:1804.02767, 2018.

[6] Cui Y, Chen R, Chu W, et al. Deep learning for image and point cloud fusion in autonomous driving: A review. IEEE Transactions on Intelligent Transportation Systems. 2021, 23(2): 722-739.

[7] Bai X, Hu Z, Zhu X, et al. Transfusion: Robust lidar-camera fusion for 3d object detection with transformers. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022, 1090-1099.

[8] Wang C Y, Yeh I H, Mark Liao H Y. Yolov9: Learning what you want to learn using programmable gradient information. European conference on computer vision. Cham: Springer Nature Switzerland. 2024, 1-21.

[9] Ke L, Danelljan M, Li X, et al. Mask Transfiner for High-Quality Instance Segmentation. arXiv, 2021(2021).

[10] Cao C, Wang B, Zhang W, et al. An improved faster R-CNN for small object detection. Ieee Access. 2019, 7: 106838-106846.

[11] Huangfu JY, Meng Q, Meng LC, Xie YP. YOLOv5 Traffic Object Detection Based on GhostNet and Attention Mechanism. Computer Systems and Applications, 2023, 32(4): 149-160(in Chinese).

[12] Zhao Z, Kang J, Wu B, et al. AE-Net: a high accuracy and efficient network for railway obstacle detection based on convolution and transformer. IEEE Transactions on Instrumentation and Measurement. 2024, 73: 1-14.

[13] Li R, He C, Li S, et al. DynaMask: dynamic mask selection for instance segmentation. roceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023, 11279-11288.

[14] Wang W, Dai J, Chen Z, et al. Internimage: Exploring large-scale vision foundation models with deformable convolutions. Proceedings of the IEEE/CVF conference on computer vision

ISSN 2959-6157

and pattern recognition. 2023, 14408-14419.
[15] Wei Z, Liu L, Hu Q X, et al. Methods, devices, equipment and media for detecting and positioning of railway unmanned driving obstacles: CN202510596776.2. CN120107933A [2025-09-26].
[16] Tagiew R, Klasek P, Tilly R, et al. Osdar23: Open sensor data for rail 2023. 2023 8th International Conference on Robotics and Automation Engineering (ICRAE). IEEE, 2023, 270-276.
[17] Diaz X, D'Amico G, Dominguez-Sanchez R, et al. Towards Railway Domain Adaptation for LiDAR-based 3D Detection: Road-to-Rail and Sim-to-Real via SynDRA-BBox. arxiv preprint arxiv:2507, 16413, 2025.
[18] Li X, Peng X. Rail detection: An efficient row-based network and a new benchmark. Proceedings of the 30th ACM International Conference on Multimedia. 2022, 6455-6463.
[19] Zhou D W, Zhang Y, Wang Y, et al. Learning without forgetting for vision-language models. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2025.